Journal of Science and Social Research ISSN 2615 – 4307 (Print)
June 2026, IX (3): 5159 – 5165 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

SEGMENTASI JAMAAH UMRAH MENGGUNAKAN K-MEANS DENGAN METODE ELBOW GUNA STRATEGI PENINGKATAN MANAJEMEN SUMBER DAYA MANUSIA

Muhammad Iqbal[1], Nirda Julianda[2], Nadin Andira[2], Aleeya Dharani[1]
[1)]Sistem Informasi, Universitas Royal
[2)]Manajemen, Universitas Royal
Email: [1]iqbalmh@royal.ac.id

Abstract: *The increasing number of Umrah pilgrims requires travel agencies to implement data-driven approaches to improve service quality and support Human Resource Management (HRM). This study aims to segment Umrah pilgrims using the K-Means clustering algorithm as a basis for HRM strategies. The research follows the Knowledge Discovery in Databases (KDD) process, including data preprocessing, clustering in RapidMiner, and cluster evaluation using the Elbow Method based on performance distance (average within centroid distance). Euclidean Distance was used to measure similarity among data objects. Seven clustering experiments were conducted with k = 2–8, producing performance distance values of 3418.971, 1416.677, 865.316, 479.795, 344.203, 258.804, and 214.562, respectively. The Elbow curve indicates that k = 3 is the optimal number of clusters because it represents the most significant decrease before the curve stabilizes. The resulting clusters provide objective information for supporting staff placement, service task allocation, pilgrim assistance planning, and employee competency development. Therefore, the integration of K-Means and the Elbow Method offers an effective data-driven approach for supporting HRM decision-making in Umrah travel agencies.*

Keywords: *Elbow Method; Human Resource Management; K-Means Clustering; Performance Distance; Umrah Pilgrims.*

Abstrak: Peningkatan jumlah jamaah umrah mendorong biro perjalanan memanfaatkan analisis data untuk meningkatkan kualitas pelayanan dan mendukung pengambilan keputusan pada Manajemen Sumber Daya Manusia (MSDM). Penelitian ini bertujuan melakukan segmentasi jamaah umrah menggunakan algoritma K-Means Clustering sebagai dasar penyusunan strategi MSDM. Metode penelitian mengikuti tahapan Knowledge Discovery in Databases (KDD), meliputi data preprocessing, proses klasterisasi pada RapidMiner, serta evaluasi menggunakan Metode Elbow berdasarkan performance distance (average within centroid distance). Pengukuran kemiripan data dilakukan menggunakan Euclidean Distance. Pengujian dilakukan sebanyak tujuh kali dengan variasi k = 2–8, menghasilkan nilai performance distance berturut-turut 3418,971; 1416,677; 865,316; 479,795; 344,203; 258,804; dan 214,562. Hasil evaluasi menunjukkan bahwa k = 3 merupakan jumlah klaster optimal karena membentuk titik siku (elbow point) dengan penurunan nilai paling signifikan sebelum kurva melandai. Hasil segmentasi dapat dimanfaatkan sebagai dasar penempatan pegawai, pembagian tugas pelayanan, penyusunan tim pendamping jamaah, dan pengembangan kompetensi pegawai secara lebih tepat sasaran.

Kata kunci: Jamaah Umrah; K-Means Clustering; Manajemen Sumber Daya Manusia; Metode Elbow; Performance Distance.

PENDAHULUAN

Berdasarkan laporan Kementerian Agama RI yang dikutip dalam artikel ilmiah tahun 2025, jumlah jamaah umrah Indonesia meningkat dari 957.016 orang (2022) menjadi 1.368.616 orang (2023) dan diperkirakan mencapai 1,9 juta orang

pada 2024. Peningkatan ini menunjukkan bahwa penyelenggara perjalanan ibadah umrah (PPIU) harus mengelola data jamaah dalam jumlah yang semakin besar (Nahendra, 2025). Data tersebut tidak hanya berisi informasi administratif, tetapi juga mencerminkan karakteristik jamaah, seperti usia, jenis kelamin, asal daerah, pilihan paket perjalanan, dan jadwal keberangkatan. Apabila dianalisis secara tepat, data tersebut dapat menghasilkan informasi yang bernilai untuk meningkatkan kualitas pelayanan, mendukung pengambilan keputusan, serta mengoptimalkan pengelolaan sumber daya manusia pada biro perjalanan umrah. Namun, pada praktiknya sebagian besar data jamaah masih dimanfaatkan sebagai arsip administrasi sehingga potensi informasi yang terkandung di dalamnya belum dimanfaatkan secara optimal (Salminen et al., 2023). Salah satu pendekatan yang banyak digunakan untuk mengekstraksi pengetahuan dari kumpulan data berukuran besar adalah data mining.

Dalam konteks machine learning, teknik analisis data secara umum dibedakan menjadi supervised learning dan unsupervised learning. Pendekatan supervised learning memanfaatkan data berlabel untuk membangun model prediksi atau klasifikasi, sedangkan unsupervised learning digunakan untuk menemukan pola atau struktur yang tersembunyi pada data tanpa label. Salah satu teknik yang paling banyak digunakan pada pendekatan unsupervised learning adalah clustering, yaitu proses mengelompokkan data berdasarkan tingkat kemiripan karakteristik antarobjek(Bishop & Bishop, 2024; James et al., 2023). Salah satu teknik unsupervised learning yang paling banyak diterapkan adalah clustering, yaitu proses mengelompokkan objek berdasarkan tingkat kemiripan karakteristik sehingga objek dalam satu kelompok memiliki tingkat homogenitas yang tinggi dibandingkan dengan kelompok lainnya (Salminen et al., 2023).

Pada bidang penyelenggaraan perjalanan ibadah umrah, penelitian mengenai penerapan K-Means Clustering masih relatif terbatas. Salah satu penelitian yang dilakukan oleh (Iqbal, 2019) berhasil mengelompokkan data jamaah umrah berdasarkan atribut usia, jenis kelamin, dan paket perjalanan sehingga diperoleh tiga kelompok jamaah yang dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran biro perjalanan. Meskipun demikian, penelitian tersebut masih berfokus pada aspek pemasaran dan belum mengaitkan hasil segmentasi dengan pengambilan keputusan pada aspek Manajemen Sumber Daya Manusia (MSDM), seperti penempatan pegawai, pembagian tugas pelayanan, penyusunan tim pendamping jamaah, maupun pengembangan kompetensi pegawai.

Teknik machine learning secara umum dikelompokkan menjadi dua pendekatan utama, yaitu supervised learning dan unsupervised learning. Supervised learning digunakan untuk membangun model prediksi berdasarkan data yang telah memiliki label, sedangkan unsupervised learning bertujuan menemukan pola atau struktur tersembunyi dari data tanpa memerlukan pelabelan sebelumnya. Pemilihan pendekatan bergantung pada karakteristik data dan tujuan analisis yang ingin dicapai. Pada data yang belum memiliki label, pendekatan unsupervised learning lebih sesuai karena mampu mengidentifikasi kemiripan antarobjek secara otomatis (Sáez-Ortuño et al., 2023)

Tujuan penelitian ini adalah menerapkan algoritma K-Means Clustering untuk melakukan segmentasi karakteristik jamaah umrah berdasarkan tingkat kemiripan atribut sehingga diperoleh kelompok jamaah yang objektif dan representatif. Untuk meningkatkan kualitas hasil pengelompokan, penelitian ini mengintegrasikan Metode Elbow dengan indikator performance distance (average within centroid distance) dalam menentukan jumlah klaster yang paling optimal. Pendekatan ini dipilih karena K-Means merupakan salah satu algoritma clustering yang memiliki efisiensi komputasi tinggi dan banyak digunakan

dalam penelitian segmentasi pelanggan maupun pengambilan keputusan berbasis data (John et al., 2023; Salminen et al., 2023). Seluruh proses analisis dan pemodelan dilakukan menggunakan perangkat lunak RapidMiner, yang menyediakan lingkungan visual untuk implementasi algoritma clustering dan evaluasi performa klaster secara sistematis (RapidMiner GmbH, 2024). Hasil segmentasi yang diperoleh diharapkan tidak hanya memberikan

gambaran karakteristik jamaah secara objektif, tetapi juga menjadi dasar dalam mendukung strategi Manajemen Sumber

Daya Manusia (MSDM) pada biro perjalanan umrah, meliputi penempatan pegawai, pembagian tugas pelayanan, penyusunan tim pendamping jamaah, serta pengembangan kompetensi pegawai sesuai karakteristik masing-masing segmen jamaah.

METODE

Untuk menyiapkan data dari file dataset.xlsx agar siap digunakan dalam algoritma K-Means Clustering, diperlukan prapemrosesan (preprocessing) data, khususnya Normalisasi Data.

Analisis Awal Data

Berdasarkan pemeriksaan file dataset.xlsx, terdapat 5 kolom utama: NO., NAMA, SEX, USIA, dan PAKET.
Kolom Non-Numerik / Identitas (NO., NAMA): Kolom ini harus dihapus/diabaikan saat proses clustering karena tidak memiliki nilai matematis untuk perhitungan jarak antar-cluster. Kolom Numerik (SEX, USIA, PAKET): Kolom inilah yang akan digunakan sebagai fitur clustering. Namun, terdapat perbedaan skala yang sangat mencolok: SEX memiliki rentang 0 hingga 1. USIA memiliki rentang 2 hingga 84 tahun (rata-rata 53 tahun). PAKET memiliki rentang 10 hingga 13. Ini akan menjadi masalah untuk K-Means karena K-Means menggunakan perhitungan Jarak Euclidean untuk menentukan kemiripan fitur dengan skala besar (USIA) akan mendominasi perhitungan jarak secara mutlak, sehingga fitur SEX dan PAKET menjadi tidak berpengaruh (Wongoutong, 2024).

Tahapan Normalisasi Data yang Digunakan

Metode yang digunakan untuk kasus ini adalah Min-Max Normalization. Dengan formula Rumus Min-Max Normalization:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

antar-data. Jika data tidak dinormalisasi,

Tujuan penggunaan Min-Max tersebut untuk menyamakan Skala ke Rentang [0, 1]: Semua fitur (SEX, USIA, PAKET) akan diubah skalanya menjadi tepat berada di antara 0 dan 1. Dengan begitu, setiap fitur memiliki bobot (weight) yang adil dalam perhitungan jarak K-Means. Mempertahankan Proporsi Jarak Asli, Dimana Min-Max normalization tidak mengubah distribusi asli data, melainkan hanya menyusutkan skalanya secara proporsional. Sesuai untuk Fitur Kategorikal yang Sudah di-Encode: Kolom SEX (0 dan 1) tidak akan terganggu nilainya karena batas minimumnya sudah 0 dan maksimumnya 1, sementara USIA dan PAKET akan menyesuaikan. Dataset sebelum dilakukan Normalisasi.

Tabel 1. Dataset Sebelum Normalisasi

No.	Nama	Sex	Usia	Paket
1	IMAM …	1	72	10
2	DAMRO …	0	55	10
116	HERNAWATI …	0	62	13
117	LASMEN …	1	60	13
258	MARIA …	0	50	12
259	SUKARDI …	1	53	12
260	FRENDIKA …	1	15	12
261	NURSALATI ..	0	58	10
262	ISMANTO …	1	65	10
263	SUHARNI …	0	61	10

Tabel 2. Dataset Setelah Normalisasi

No.	Nama	Sex	Usia	Paket
1	IMAM …	1	0.8537	0
2	DAMRO …	0	0.6463	0
116	HERNAWATI …	0	0.7317	1
117	LASMEN …	1	0.7073	1
258	MARIA …	0	0.5854	0.6667
259	SUKARDI …	1	0.622	0.6667
260	FRENDIKA …	1	0.1585	0.6667
261	NURSALATI ..	0	0.6829	0
262	ISMANTO …	1	0.7683	0
263	SUHARNI …	0	0.7195	0

Metodologi penelitian ini dirancang secara sistematis agar proses segmentasi data dapat direplikasi oleh peneliti lain dengan prosedur yang sama. Penelitian menggunakan dataset jamaah umrah yang belum memiliki label (unlabeled data), sehingga pendekatan unsupervised learning dipilih sebagai metode analisis. Dataset yang digunakan terdiri atas atribut-atribut yang merepresentasikan karakteristik jamaah, seperti usia, jenis kelamin, asal daerah, jenis paket perjalanan, dan atribut lain yang relevan. Sebelum proses klasterisasi dilakukan, data melalui tahapan preprocessing yang meliputi data cleaning, transformasi data, dan normalisasi untuk meningkatkan kualitas data serta mengurangi pengaruh perbedaan skala antaratribut terhadap hasil pengelompokan (Wongoutong, 2024).

Proses klasterisasi dilakukan menggunakan algoritma K-Means Clustering yang diimplementasikan melalui perangkat lunak RapidMiner Studio. Perangkat lunak ini dipilih karena menyediakan lingkungan analisis visual yang mendukung implementasi algoritma machine learning, pengelolaan alur kerja (workflow), serta evaluasi hasil klaster secara sistematis (RapidMiner GmbH, 2024). Penelitian melakukan beberapa skenario pengujian dengan variasi jumlah klaster (k = 2 hingga k = 7) untuk memperoleh struktur klaster yang paling representatif terhadap karakteristik data jamaah.

Prinsip dasar algoritma K-Means adalah mengelompokkan objek berdasarkan tingkat kemiripan yang diukur menggunakan Euclidean Distance. Setiap objek akan ditempatkan pada klaster dengan centroid terdekat, kemudian posisi centroid diperbarui secara iteratif berdasarkan nilai rata-rata seluruh anggota klaster hingga tercapai kondisi konvergen, yaitu ketika perubahan centroid tidak lagi signifikan (John et al., 2023). Karena perhitungan Euclidean Distance sangat dipengaruhi oleh rentang nilai setiap atribut, penelitian ini menerapkan proses normalisasi agar seluruh atribut memberikan kontribusi yang seimbang terhadap pembentukan klaster (Wongoutong, 2024)

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

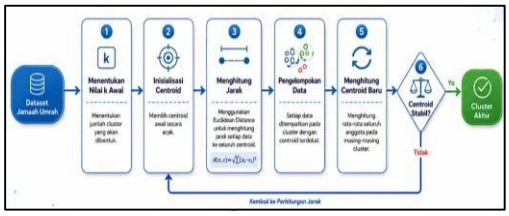
Gambar 1. Persamaan untuk Pengukuran Jarak Euclidean

Di mana d melambangkan jarak geometris antara dua poin data, x_i menyatakan nilai atribut ke-i dari data pertama, y_i menyatakan nilai atribut ke-i dari data kedua, dan n merupakan total jumlah atribut yang dikomparasi di dalam dataset.

Algoritma K-Means Clustering merupakan algoritma partisi yang bekerja secara iteratif untuk mengelompokkan objek berdasarkan kedekatannya terhadap pusat klaster (centroid). Proses diawali dengan menentukan jumlah klaster (k) yang akan dibentuk, kemudian memilih posisi awal centroid sebagai titik awal proses klasterisasi. Selanjutnya, setiap data dihitung jaraknya terhadap seluruh centroid menggunakan Euclidean Distance, kemudian setiap objek dialokasikan ke klaster yang memiliki centroid terdekat. Setelah seluruh objek memperoleh keanggotaan klaster, centroid

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
June 2026, IX (3): 5159 – 5165 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

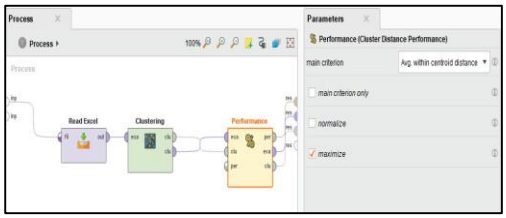
diperbarui dengan menghitung nilai rata-rata (mean) dari seluruh anggota pada masing-masing klaster. Proses penghitungan jarak, pembaruan centroid, dan pengalokasian ulang anggota klaster dilakukan secara berulang hingga posisi centroid tidak lagi mengalami perubahan yang signifikan atau telah memenuhi kriteria konvergensi. Mekanisme iteratif tersebut menjadikan K-Means sebagai salah satu algoritma clustering yang efisien dan banyak digunakan untuk analisis data berukuran besar karena memiliki kompleksitas komputasi yang relatif rendah serta mudah diimplementasikan (Nazeer & others, 2023)



Clustering

Setelah seluruh data teralokasi ke dalam klaster masing-masing, posisi centroid diperbarui dengan menghitung nilai rata-rata (mean) dari seluruh anggota pada setiap klaster. Proses perhitungan jarak, pengelompokan data, dan pembaruan centroid dilakukan secara berulang hingga posisi centroid tidak lagi mengalami perubahan yang signifikan atau telah mencapai kondisi konvergen (converged). Tahapan kerja algoritma tersebut mengacu pada penelitian Wu (2012).

Untuk mengevaluasi kualitas hasil pengelompokan, penelitian ini memanfaatkan **Performance Operator** pada perangkat lunak RapidMiner untuk menghitung nilai performance distance (average within-cluster distance), yaitu rata-rata jarak setiap data terhadap centroid klaster tempat data tersebut berada. Nilai tersebut digunakan sebagai indikator kualitas hasil clustering, di mana semakin kecil nilai performance distance, semakin baik tingkat homogenitas anggota dalam suatu klaster.



Gambar 3. Proses Implementasi Algoritma K-Means dan Evaluasi Cluster Distance Performance pada RapidMiner

Penentuan jumlah klaster yang optimal dilakukan menggunakan Metode Elbow sebagaimana dijelaskan oleh Bramer (2016). Metode ini bekerja dengan membandingkan nilai performance distance yang dihasilkan dari beberapa variasi jumlah klaster (k). Selanjutnya, nilai-nilai tersebut divisualisasikan dalam bentuk grafik untuk mengidentifikasi titik siku (elbow point), yaitu titik ketika penurunan nilai performance distance mulai melandai. Titik tersebut menunjukkan bahwa penambahan jumlah klaster setelahnya tidak lagi memberikan peningkatan kualitas clustering yang signifikan sehingga dipilih sebagai jumlah klaster yang paling optimal.

Pada penelitian ini dilakukan pengujian terhadap variasi jumlah klaster k = 2 hingga k = 7. Berdasarkan hasil evaluasi menggunakan performance distance, diperoleh bahwa k = 3 merupakan jumlah klaster yang paling optimal karena menghasilkan pola elbow point dengan penurunan nilai rata-rata jarak yang paling signifikan dibandingkan variasi jumlah klaster lainnya. Konfigurasi parameter utama yang digunakan pada operator K-Means di RapidMiner disajikan pada Tabel 1.

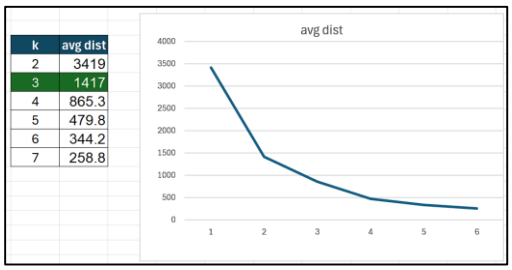
Tabel 1. Parameter Utama Operator K-Means RapidMiner

Atribut Parameter	Fungsi Komputasi	Parameter / Nilai Pengujian
k	Menentukan jumlah kelompok/klaster jamaah yang akan dibentuk	Variasi nilai 2 sampai 7
Max Runs	Jumlah iterasi	10

	maksimum proses klasterisasi untuk hasil terbaik	
Distance Measure	Metode pengukuran jarak antar data atribut	Mix Mazer / Euclidean
Max Optimization Step	Batas maksimum perpindahan posisi centroid sebelum iterasi dihentikan	100

HASIL DAN PEMBAHASAN

Proses eksekusi data jamaah umroh menggunakan RapidMiner memberikan hasil pembagian klaster yang bervariasi bergantung pada nilai parameter *k* yang dimasukkan. Berdasarkan hasil pengujian performa kuadratik jarak dalam klaster (within cluster distance) dari nilai *k*=2 hingga *k*=7, didapatkan pola penurunan nilai yang diplot menjadi sebuah grafik evaluasi. Penentuan nilai *k* optimal secara objektif dianalisis melalui Elbow Method dengan melihat visualisasi grafik yang terbentuk (Bramer, 2016). Grafik evaluasi performa Elbow Method disajikan pada Gambar 1 di bawah ini:



Gambar 4. Grafik Elbow Method untuk Penentuan K Optimal

Berdasarkan visualisasi grafik pada Gambar 1, terlihat terjadinya penurunan nilai within cluster distance yang signifikan saat mengubah nilai *k* dari 2 menjadi 3. Namun, pada perpindahan nilai *k* berikutnya (dari 4 hingga 7), penurunan nilai jarak rata-rata mulai melambat dan cenderung melandai. Titik tekukan atau "siku" grafik terlihat sangat jelas berada pada koordinat *k* =3. Dengan demikian, berdasarkan landasan teoritis Elbow Method, pembagian kelompok terbaik dan paling optimal untuk karakteristik jamaah umroh pada penelitian ini adalah berjumlah 3 klaster (*k*=3).

Setelah menetapkan nilai k=3, RapidMiner mendistribusikan data objek jamaah ke dalam tiga klaster dengan sebaran yang tidak merata, yang dipengaruhi oleh kedekatan jarak fisik atribut masing-masing jamaah. Klaster 0 mendominasi dengan jumlah data terbesar (117 item), diikuti oleh Klaster 1 (111 items) dan Klaster 2 (82 items) yang memiliki jumlah anggota lebih sedikit. Hasil klaster menunjukkan karakteristik bahwa klaster 1 memiliki karakteristik Jamaah Usia Lanjut, klaster 2 dengan jamaah usia produktif dan klaster 3 jamaah VIP/Premium.

SIMPULAN

Hasil segmentasi penelitian terdiri tiga kelompok jamaah dengan karakteristik yang berbeda sehingga dapat memberikan gambaran yang lebih terstruktur mengenai profil jamaah. Informasi hasil segmentasi jamaah dapat dimanfaatkan sebagai bahan pertimbangan oleh manajemen dalam mengatur penugasan pendamping jamaah, merencanakan kebutuhan personel pada setiap kelompok keberangkatan, serta menyusun pola pelayanan yang lebih sesuai dengan karakteristik jamaah.

DAFTAR PUSTAKA

Bishop, C. M., & Bishop, H. (2024). *Pattern Recognition and Machine Learning* (2nd ed.). Springer. https://doi.org/10.1007/978-3-031-45468-4

Iqbal, M. (2019). Klasterisasi Data Jamaah Umroh pada Auliya Tour \&

Travel Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, *6*(1), 45–52. https://jurnal.stmikroyal.ac.id/index.php/jurteksi/article/view/352

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python* (2nd ed.). Springer. https://doi.org/10.1007/978-3-031-38747-0

John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, *2*(4), 809–823. https://doi.org/10.3390/analytics2040042

Nahendra, et A. (2025). The Role of Public Trust in Moderating the Influence of Service Quality on Public Satisfaction. *INDONESIAN JOURNAL OF ECONOMICS, SOCIAL, AND HUMANITIES*, *7*(7), 827–833. https://doi.org/10.35629/5252-0707827833

Nazeer, K. A. A., & others. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

RapidMiner GmbH. (2024). *RapidMiner Documentation*. https://docs.rapidminer.com/latest/

Sáez-Ortuño, L., Huertas-Garcia, R., Forgas-Coll, S., & Puertas-Prats, E. (2023). How can entrepreneurs improve digital market segmentation? A comparative analysis of supervised and unsupervised learning algorithms. *International Entrepreneurship and Management Journal*, *19*(4), 1893–1920. https://doi.org/10.1007/s11365-023-00882-1

Salminen, J., Mustak, M., Sufyan, M., & Jansen, B. J. (2023). How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation. *Journal of Marketing Analytics*, *11*(4), 677–692. https://doi.org/10.1057/s41270-023-00235-5

Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLoS ONE*, *19*(12), 1–19. https://doi.org/10.1371/journal.pone.0310839